

A Review: Generative Artificial Intelligence

Ms. Kusum Yadav¹, Tushar Agarwal¹, Vedansh Sharma³,

¹Assistant Professor Department of Information Technology, Jaipur Engineering College & Research Centre, Jaipur

²Student Department of Information Technology, Jaipur Engineering College and Research Centre, Jaipur

Email: tusharagarwal.it25@jecrc.ac.in

³Student Department of Information Technology, Jaipur Engineering College and Research Centre, India

Email : vedanshsharma.it25@gmail.com

Abstract— Generative Artificial Intelligence (GenAI) is increasingly being adopted across diverse fields, from industrial applications to academic research. These systems are primarily operated through prompts— user-provided instructions that guide the AI in performing specific tasks. However, due to the field's recent emergence, prompt engineering still lacks standardized terminology and clearly defined boundaries. This paper aims to bring structure to this evolving domain by proposing a classification of different prompting strategies and analyzing their real-world applications. It offers a comprehensive glossary of frequently used terms, a well-organized framework encompassing various prompting techniques across different input types, and best practices for working with large language models (LLMs) such as ChatGPT. The study also includes a meta-analysis of existing literature related to text-based prompting, positioning this work as a foundational reference in the area.

Keywords—About five keywords in alphabetical order, separated by comma

I. INTRODUCTION

Transformer-based LLMs have become integral to various applications, ranging from personal use cases and enterprise tools to academic research. These models typically operate by responding to user inputs known as prompts, which may be in the form of questions, commands, or other instructions. While prompts are most often text-based, they can also include multimedia input such as images, audio, and video. One of the key strengths of LLMs is their ability

to understand and respond to natural language prompts, which contributes to their adaptability and broad applicability. The effectiveness of these models often hinges on the quality and structure of the prompts used. Well-crafted prompts generally

Despite its growing relevance, prompt engineering remains a relatively nascent discipline with ongoing debates and inconsistencies in how techniques are defined and applied. To address this, our study undertakes a detailed review of the current methods and terminology associated with prompt engineering. The goal is to offer a clearer, more systematic understanding of the field by developing a shared vocabulary and structured categorization of prompting approaches.

What is a Prompt?—A prompt refers to any type of input provided to a generative AI system to direct its response. This input can take the form of text, images, audio, or other data types. For example, a prompt might be a request like “Compose a three-paragraph email for an accounting firm’s marketing campaign,” an image with a question overlaid, or an audio clip with instructions like “Summarize this meeting.” While most current prompts are text-based, this could change as multimodal models capable of interpreting multiple types of media become more prevalent.

To simplify prompt generation, developers often use prompt templates—predefined structures where placeholders are filled with specific content to form complete prompts. For instance, a template might say: “Write a poem about the following topic:

{USER_INPUT},” where {USER_INPUT} can be dynamically replaced with different themes. This

makes prompt design more efficient and scalable for repeated tasks.

A Meta-Analysis of Prompting Systematic Review Process

To develop a credible and high-quality dataset of academic resources for this study, we implemented a systematic literature review strategy inspired by PRISMA guidelines. This structured approach involved multiple stages to ensure a consistent and comprehensive selection of relevant studies. The resulting dataset, along with detailed metadata documentation, is publicly accessible via HuggingFace.

We gathered research papers primarily from databases such as arXiv, Semantic Scholar, and the In-Context Learning (ICL) In-Context Learning refers to the capacity of generative AI systems to perform tasks by interpreting examples or instructions embedded within the

prompt itself—without requiring model retraining or weight adjustment (Brown et al., 2020; Radford et al., 2019b). These tasks may be inferred through demonstrations (Figure 2.4) or directives (Figure 2.5). However, the term "learning" can be misleading in this context; ICL often simply involves task specification rather than the acquisition of new skills—many capabilities may already exist in the model's training data (Figure 2.6). For more detail, see Appendix A.9. There is ongoing research into refining and better understanding ICL (Bansal et al., 2023).

Thought Generation

This approach involves prompting language models to explicitly articulate their reasoning while tackling a problem (Zhang et al., 2023c). Chain-of-Thought (CoT) prompting, introduced by Wei et al. (2022b), uses a few-shot method to encourage models to reveal their logical process prior to giving a final answer. This method, also known as Chain-of-Thoughts (Tutunov et al., 2023; Besta et al., 2024; Chen et al., 2023d), has proven effective for enhancing performance on reasoning and mathematical tasks. In the original demonstration (Wei et al., 2022b), the prompt includes a sample with a question, a reasoning path, and the correct response.

Decomposition Techniques

Researchers have extensively explored the strategy of breaking complex problems into simpler sub-tasks—a method that mirrors human problem-solving and benefits GenAI systems as well (Pate et al., 2022). While CoT often results in implicit decomposition, explicitly fragmenting a task can further enhance a model's effectiveness.

Least-to-Most Prompting

Zhou et al. (2022a) proposed a method that begins by instructing the model to deconstruct a complex task into smaller, manageable sub-tasks without solving them immediately. These sub-tasks are then addressed one at a time, with each solution appended to the prompt before moving on to the next. This approach has shown marked improvements in tasks involving logic, symbolic reasoning, and compositional generalization.

Decomposed Prompting (DECOMP)

Khot et al. (2022) introduced DECOMP, a few-shot prompting method that demonstrates the use of specific tools or functions—such as string

manipulation or web search—by splitting problems into sub-tasks delegated across different AI operations. This modular strategy outperformed Least-to-Most prompting in various scenarios, emphasizing the benefits of functional task decomposition.

Plan-and-Solve Prompting

Wang et al. (2023f) advanced this idea with a refined Zero-Shot Chain-of-Thought prompt: "Let's first understand the problem and devise a plan to solve it. Then, let's carry out the plan and solve the problem step by step." This method encourages clearer, more structured reasoning and has outperformed traditional Zero-Shot CoT approaches on several benchmark reasoning tasks.

Tree-of-Thought (ToT) (Yao et al., 2023b), also known as Tree of Thoughts (Long, 2023), creates a tree-like search problem by starting with an initial problem then generating multiple possible steps in the form of thoughts (as from a CoT). It evaluates the progress each step makes towards solving the problem (through prompting) and decides which steps to continue with, then keeps creating more

thoughts.

To Tis particularly effective for tasks that require search and planning.

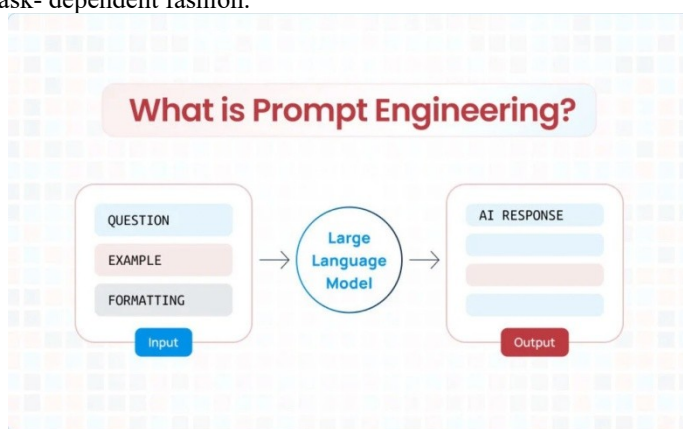
Recursion-of-Thought (Lee and Kim, 2023) is similar

to regular CoT. However, every time it encounters a complicated problem in the middle of its reasoning chain, it sends this problem into another prompt/LLM call. After this is completed, the answer is inserted into the original prompt. In this way, it can recursively solve complex problems, including ones which might otherwise run over that maximum context length.

This

method has shown improvements on arithmetic and algorithmic tasks. Though implemented using fine-tuning to output a special token that sends sub-problem into another prompt, it could also be done only through prompting. Program-of-Thoughts (Chen et al., 2023d) uses LLMs like Codex to generate programming code as reasoning steps. A code interpreter executes these steps to obtain the final answer. It excels in mathematical and programming-related tasks but is less effective for semantic reasoning tasks.

Faithful Chain-of-Thought (Lyu et al., 2023) generates a CoT that has both natural language and symbolic language (e.g. Python) reasoning, just like Program-of-Thoughts. However, it also makes use of different types of symbolic languages in a task-dependent fashion.



II.

CONCLUSION

answer. It excels in mathematical and programming-related tasks but is less effective for semantic reasoning tasks.

Faithful Chain-of-Thought (Lyu et al., 2023) generates a CoT that has both natural language and symbolic language (e.g. Python) reasoning, just like Program-of-Thoughts. However, it also makes use of different types of symbolic languages in a task-dependent fashion.

meetings continue to be organized. However, during this process, it was easier and more comfortable for some people to work online; therefore, their interest in Metaverse has increased considerably. It is a matter of curiosity whether people will accustom themselves to such a virtual environment that offers positive and negative results. In addition to psychological issues followed by the Metaverse, it is also discussed that some physical problems may occur on the way. According to Kuş (2021), users may use some virtual reality headsets to move in the Metaverse virtual world, resulting in some actions that can harm themselves in the physical world. This raises the possibility that Metaverse could be a physical disadvantage, too. Until a few years ago, the advertising industry differed from the current one. Brands used to promote their products through television and radio, now they can promote them on the internet and reach their users easily and at lower costs. Since people spend most of their time online, it makes more sense to advertise and promote products online. As Bilgici and Şisman (2022) mentioned in their study, social media advertising will now shift from social media channels such as Twitter, Instagram, etc. to the Metaverse world with the emergence of the concept of 'metafluencer' and brand strategies will now start to be built on 'metafluencer' people. The concept of the Metaverse, both with its positive and negative aspects, is slowly entering human lives and significantly affecting their traditional lifestyle you are working with constitute a good representation of that problem. It is better to start

with simpler approaches first, and to remain skeptical of claims about method performance. To those already engaged in prompt engineering, we hope that our taxonomy will shed light on the relationships

REFERENCES

1. Adept. 2023. ACT-1: Transformer for Actions. <https://www.adept.ai/blog/act-1>. Rishabh Agarwal, Avi Singh, Lei M Zhang, Bernd Bohnet, Luis Rosias, Stephanie Chan, Biao Zhang, Ankesh Anand, Zaheer Abbas, Azade Nova, et al. 2024. Many-shot in-context learning. arXiv preprint arXiv:2404.11018. Sweta Agrawal, Chunting Zhou, Mike Lewis, Luke Zettlemoyer, and Marjan Ghazvininejad. 2023. In-context examples selection for machine translation. In Findings of the Association for Computational Linguistics: ACL 2023, pages 8857–8873, Toronto, Canada. Association for Computational Linguistics. Kabir Ahuja, Harshita Diddee, Rishav Hada, Millicent Ochieng, Krithika Ramesh, Prachi Jain, Akshay Nambi, Tanuja Ganu, Sameer Segal, Maxamed Axmed, Kalika Bali, and Sunayana Sitaram. 2023. MEGA: Multilingual Evaluation of Generative AI. In EMNLP. Rebuff AI. 2023. A self-hardening prompt injection detector. Anirudh Ajith, Chris Pan, Mengzhou Xia, Ameet Deshpande, and Karthik Narasimhan. 2024. InstructEval: Systematic evaluation of instruction selection methods. In Findings of the Association for Computational Linguistics: NAACL 2024, pages 4336–4350, Mexico City, Mexico. Association for Computational Linguistics. Silvia Araújo and Micaela Aguiar. 2023. Comparing chatgpt’s and human evaluation of scientific texts’ translations from english to portuguese using popular automated translators. CLEF. Arthur AI. 2024. Arthur shield. Akari Asai, Sneha Kudugunta, Xinyan Velocity Yu, Terra Blevins, Hila Gonen, Machel Reid, Yulia Tsvetkov, Sebastian Ruder, and Hannaneh Hajishirzi. 2023. BUFFET: Benchmarking Large Language Models for Few-shot Cross-lingual Transfer. Muhammad Awais, Muzammal Naseer, Salman Khan, Rao Muhammad Anwer, Hisham Cholakkal, Mubarak Shah, Ming-Hsuan Yang, and Fahad Shahbaz Khan. 2023. Foundational models defining a new era in vision: A survey and outlook. Abhijeet Awasthi, Nitish Gupta, Bidisha Samanta, Shachi Dave, Sunita Sarawagi, and Partha Talukdar. 2023. Bootstrapping multilingual semantic parsers using large language models. In Proceedings of the between existing techniques. To those developing new techniques, we encourage situating new methods within our taxonomy, as well as including ecologically valid case studies and illustrations of those techniques. 17th Conference of the European Chapter of the Association for Computational Linguistics, pages 2455–2467, Dubrovnik, Croatia.